Aneuploidy-Dependent Massive Deregulation of the Cellular Transcriptome and Apparent Divergence of the Wnt/β-catenin Signaling Pathway in Human Rectal Carcinomas

Marian Grade,[1,3] B. Michael Ghadimi,[1] Sudhir Varma,[2] Richard Simon,[2] Danny Wangsa,[3] Linda Barenboim-Stapleton,[3] Torsten Liersch,[1] Heinz Becker,[1] Thomas Ried,[3] and Michael J. Difilippantonio[3]

[1]Department of General Surgery, University Medical Center, Göttingen, Germany and [2]Biometrics Research Branch and [3]Genetics Branch, National Cancer Institute, NIH, Bethesda, Maryland

Abstract

To identify genetic alterations underlying rectal carcinogenesis, we used global gene expression profiling of a series of 17 locally advanced rectal adenocarcinomas and 20 normal rectal mucosa biopsies on oligonucleotide arrays. A total of 351 genes were differentially expressed ($P < 1.0e{-}7$) between normal rectal mucosa and rectal carcinomas, 77 genes had a >5-fold difference, and 85 genes always had at least a 2-fold change in all of the matched samples. Twelve genes satisfied all three of these criteria. Altered expression of genes such as *PTGS2* (*COX-2*), *WNT1, TGFB1, VEGF*, and *MYC* was confirmed, whereas our data for other genes, like *PPARD* and *LEF1*, were inconsistent with previous reports. In addition, we found deregulated expression of many genes whose involvement in rectal carcinogenesis has not been reported. By mapping the genomic imbalances in the tumors using comparative genomic hybridization, we could show that DNA copy number gains of recurrently aneuploid chromosome arms 7p, 8q, 13q, 18q, 20p, and 20q correlated significantly with their average chromosome arm expression profile. Taken together, our results show that both the high-level, significant transcriptional deregulation of specific genes and general modification of the average transcriptional activity of genes residing on aneuploid chromosomes coexist in rectal adenocarcinomas. (Cancer Res 2006; 66(1): 267-82)

Introduction

Colorectal cancer is the third most common cause of cancer and the second leading cause of cancer death in Europe and the United States, with ~300,000 new cases and 200,000 deaths each year (1). It is widely accepted that the malignant transformation of colorectal epithelium toward invasive carcinomas is defined by the sequential acquisition of specific genetic aberrations and tumor-specific chromosomal aneuploidies (2–5). These aneuploidies are specific for certain carcinomas, are strictly conserved, and usually emerge before the transition from premalignant lesions to invasive disease (6). It therefore seems likely that the acquisition and maintenance of these specific aneuploidies, despite ongoing genetic instability, results from a biological selection pressure. However, to date, it remains unclear to which degree these chromosomal aneuploidies affect the cellular transcriptome of rectal adenocarcinomas.

Therefore, we complemented gene expression profiling with a comprehensive mapping of genomic imbalances in a subset of patients with rectal adenocarcinomas obtained from the German Rectal Cancer Study Group Trial (7). This enabled us not only to identify genes involved in the development of rectal carcinomas, but also to evaluate the consequences of aneuploidy-associated genomic imbalances on global gene expression levels.

Materials and Methods

Patients and sample collection. The 29 patients included in this study were all treated at the Department of General Surgery, University Medical Center Göttingen, Germany, and were all participants in a multicenter, randomized prospective phase III clinical trial (CAO/ARO/AIO-94, German Rectal Cancer Study Group; ref. 7). The clinical information is summarized in Supplementary Table S1. Two biopsies were taken from representative adjacent areas of the tumor mass; one biopsy was used for histopathologic diagnosis, whereas DNA and RNA were extracted from the other. When possible, a representative biopsy of normal mucosa was also obtained at a minimal distance of 3 cm from the tumor site with care to avoid sampling distally below the anocutaneous line or proximally from the colon. Our data set of 29 carcinomas and 20 mucosa biopsies includes 12 patient-matched pairs of biopsies from tumor and normal mucosa (Supplementary Table S2).

DNA and RNA isolation. Biopsies were ascertained and stored immediately in RNAlater (Ambion, Austin, TX). Bioptic material was in the range of 5 to 80 mg. DNA and RNA extraction was done using TRIzol (Invitrogen, Carlsbad, CA) following standard procedures (http://www.riedlab.nci.nih.gov/protocols.asp).

Expression profiling. To generate enough RNA for technical replicate hybridizations, mRNA was amplified using Amino Allyl MessageAmp aRNA kit (Ambion), resulting in antisense mRNA amounts that averaged 50 μg. RNA was labeled indirectly by incorporation of aminoallyl-dUTP, followed by chemical coupling of Cy3 (Amersham, Piscataway, NJ). Control cRNA was generated by amplification of a reference mRNA pool (Stratagene, La Jolla, CA) and labeled as above using Cy5 (Amersham). RNA quantification and labeling efficiency was determined using the Nanodrop quantification device (Nanodrop, Rockland, DE).

Expression profiling was done on the National Cancer Institute oligonucleotide arrays (22,231 features) as previously described (8). Briefly, 5.0 μg of Cy3-labeled test cRNA and 5.0 μg of Cy5-labeled control cRNA were hybridized at 42°C overnight in specifically designed hybridization cassettes (TeleChem International, Sunnyvale, CA). After hybridization, slides were washed and scanned on an Axon scanner using GenePixPro (3.0) software (Axon Instruments, Union City, CA). Spot quality was assessed according to criteria in GenePixPro (3.0) software. Background subtraction and normalization was done upon extraction of the data from the Center for Information Technology/NIH microarray database, mAdb (http://nciarray.nci.nih.gov/).

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Requests for reprints: Michael J. Difilippantonio, Genetics Branch, Center for Cancer Research/National Cancer Institute/NIH, Room 1408, Building 50, 50 South Drive, Bethesda, MD 20892-8010. Phone: 301-435-3991; Fax: 301-402-1204; E-mail: difilipm@mail.nih.gov.

doi:10.1158/0008-5472.CAN-05-2533

Table 1. Expression and *P* values for genes of interest

Unigene	Gene name	Description	Map
A. Most stringently selected genes			
Hs.936	*SLC34A1*	Solute carrier family 34 (sodium phosphate), member 1	5q35
Hs.194680	*WISP1**	WNT1 inducible signaling pathway protein 1	8q24.1-q24.3
Hs.63931	*DACH**	Dachshund homologue (*Drosophila*)	13q22
Hs.413924	*TACSTD2*	Tumor-associated calcium signal transducer 2	1p32-p31
Hs.413924	*CXCL10*	Chemokine (C-X-C motif) ligand 10	4q21
Hs.97199	*C1QR1*	Complement component 1, q subcomponent, receptor 1	20p11.22
Hs.82101	*PHLDA1**	Pleckstrin homology-like domain, family A, member 1	12q15
Hs.26770	*FABP7*	Fatty acid–binding protein 7, brain	6q22-q23
Hs.58561	*GPR87*	G protein–coupled receptor 87	3q24
Hs.25960	*MYCN*	V-myc myelocytomatosis viral-related oncogene, neuroblastoma	2p24.1
Hs.444552	*FLJ12529*	Pre-mRNA cleavage factor I, 59 kDa subunit	11q12.3
Hs.194660	*CLN3*	Ceroid-lipofuscinosis, neuronal 3, juvenile	16p12.1
B. Additional genes from ingenuity pathway analysis networks in Fig. 1			
Hs.434059	*ETV4*	Ets variant gene 4 (E1A enhancer binding protein, E1AF)	17q21
Hs.437058	*STAT5A*	Signal transducer and activator of transcription 5A	17q11.2
Hs.418533	*BUB3*	BUB3 budding uninhibited by benzimidazoles 3 homologue (yeast)	10q26
Hs.411098	*NPM1*	Nucleophosmin (nucleolar phosphoprotein B23, numatrin)	5q35
Hs.156316	*DCN*	Decorin	12q13.2
Hs.76753	*ENG*	Endoglin (Osler-Rendu-Weber syndrome 1)	9q33-q34.1
Hs.202453	*MYC*	V-myc myelocytomatosis viral oncogene homologue (avian)	8q24.12-q24.13
Hs.436015	*SRC*	V-src sarcoma viral oncogene homologue (avian)	20q12-q13
Hs.93213	*BAK1*	BCL2-antagonist/killer 1	6p21.3
Hs.414795	*SERPINE1*	Serine (or cysteine) proteinase inhibitor, clade E, member 1	7q21.3-q22
Hs.9460	*SP1*	Sp1 transcription factor	12q13.1
Hs.132594	*RELA**	V-rel reticuloendotheliosis viral oncogene homologue A	11q13
Hs.141125	*CASP3*	Caspase 3, apoptosis-related cysteine protease	4q34
Hs.172609	*NUCB1**	Nucleobindin 1	19q13.2-q13.4
Hs.447905	*E2F5*	E2F transcription factor 5, p130-binding	8q21.2
Hs.389900	*BLK*	B lymphoid tyrosine kinase	8p23-p22
Hs.245188	*TIMP3*	Tissue inhibitor of metalloproteinase 3	22q12.1-q13.2
Hs.6441	*TIMP2*	Tissue inhibitor of metalloproteinase 2	17q25
Hs.408312	*TP53*	Tumor protein p53 (Li-Fraumeni syndrome)	17p13.1
C. Wnt/β-catenin pathway genes			
Hs.7957	*ADAR*	Adenosine deaminase, RNA-specific	1q21.1-q21.2
Hs.512765	*AXIN1**	Axin 1	16p13.3
Hs.226434	*BTRC*	β-transducin repeat containing	10q24.32
Hs.371468	*CCND1*	Cyclin D1 (PRAD1: parathyroid adenomatosis 1)	11q13
Hs.318381	*CSNK1A1*	Casein kinase 1, α1	5q32
Hs.446484	*CSNK2A1*	Casein kinase 2, α1 polypeptide	20p13
Hs.196083	*CTBP1**	COOH-terminal binding protein 1	4p16
Hs.196083	*CTBP1*	COOH-terminal binding protein 1	4p16
Hs.410086	*CTNNB1**	Catenin (cadherin-associated protein), β1, 88 kDa	3p21
Hs.410086	*CTNNB1*	Catenin (cadherin-associated protein), β1, 88 kDa	3p21
Hs.79306	*EIF4E*	Eukaryotic translation initiation factor 4E	4q21-q25
Hs.126057	*FRAT1*	Frequently rearranged in advanced T-cell lymphomas	10q24.2
Hs.94234	*FZD1*	Frizzled homologue 1 (*Drosophila*)	7q21
Hs.282359	*GSK3B** [†]	Glycogen synthase kinase 3β	3q13.3
Hs.290346	*MAP3K7*	Mitogen-activated protein kinase kinase kinase 7	6q16.1-q16.3
Hs.202453	*MYC*	V-myc myelocytomatosis viral oncogene homologue (avian)	8q24.12-q24.13
Hs.3532	*NLK*	Nemo-like kinase	17q11.2
Hs.106415	*PPARD* [‡]	Peroxisome proliferative activated receptor, δ	6p21.2-p21.1
Hs.124942	*PR48*	Protein phosphatase 2A 48 kDa regulatory subunit	Xp22.33
Hs.248164	*WNT1*	Wingless-type mouse mammary tumor virus (MMTV) integration site family, member 1	12q13

Table 1. Expression and *P* values for genes of interest (Cont'd)

Avg tumor/avg mucosa	*P*	Avg tumor/avg $Mucosa_0$	*P*	Avg tumor/avg $Mucosa_1$	*P*	Avg $Mucosa_1$/avg $Mucosa_0$	*P*
102.740	<1.0e−07	96.1650	5.5e−06	108.7336	2.5e−06		
13.412	<1.0e−07	16.1320	<1.0e−07	11.7794	<1.0e−07		
9.387	<1.0e−07			12.8387	<1.0e−07		
8.952	<1.0e−07	11.4070	6.7e−06	7.7406	3.0e−07		
8.944	<1.0e−07	7.2590	1.0e−07	11.0370	<1.0e−07		
5.600	<1.0e−07	5.6000	<1.0e−07	5.6000	<1.0e−07		
0.206	<1.0e−07	0.2190	4.0e−07	0.1930	<1.0e−07		
0.180	<1.0e−07	0.1740	<1.0e−07	0.1864	<1.0e−07		
0.155	<1.0e−07	0.1280	2.0e−07	0.1874	3.6e−06		
0.124	<1.0e−07	0.1220	<1.0e−07	0.1267	<1.0e−07		
0.083	<1.0e−07	0.0850	<1.0e−07	0.0810	<1.0e−07		
0.074	<1.0e−07	0.0610	<1.0e−07	0.0889	<1.0e−07		
2.7382	<1.0e−07						
0.7192	3.0e−07						
1.4868	5.0e−07						
1.9731	1.4e−06						
2.4333	2.7e−05						
1.6315	4.0e−05						
2.1110	5.1e−05						
0.5688	0.0002						
0.8047	0.0009						
1.6132	0.0066						
0.6846	0.0199						
0.8308	0.0246						
1.7159	0.0290						
0.8474	0.0376						
0.6615	0.1074						
1.1951	0.5795						
0.9517	0.5977						
1.0839	0.6481						
0.9619	0.6734						
1.4160	0.0035	1.9900	2.0e−07	1.0080	0.9166	1.9740	8.5e−06
1.2780	0.0286	1.5920	0.0005	1.0250	0.7954	1.5530	0.0055
0.9360	0.1678	0.9170	0.1788	0.9530	0.3871	0.9620	0.5700
1.3890	0.0078	1.8170	4.1e−05	1.0630	0.5896	1.7100	0.0007
1.6910	0.0418	3.7900	7.0e−07	0.7540	0.0779	5.0240	4.0e−07
0.9420	0.5369	0.9100	0.4733	0.9750	0.8021	0.9330	0.5936
0.4580	0.0046	0.1900	<1.0e−07	1.0990	0.5894	0.1730	<1.0e−07
0.9510	0.5244	0.8670	0.1698	1.0440	0.6293	0.8300	0.0466
1.2490	0.0624	1.6180	0.0605	1.3490	0.0364	2.4220	0.0005
0.9300	0.7134	1.1560	0.3366	0.6680	0.0382	0.8570	0.2728
1.4110	0.0136	2.0830	3.5e−06	0.9570	0.7100	2.1770	2.0e−07
1.1270	0.2322	1.3330	0.0146	0.9500	0.6045	1.4040	0.0197
1.4030	0.2310	3.4860	1.6e−05	0.5640	0.0026	6.1770	1.0e−07
2.0280	<1.0e−07	2.4090	<1.0e−07	1.7080	1.1e−06	1.4100	0.0086
2.8940	0.0003	7.2700	<1.0e−07	1.2630	0.0555	5.7550	3.9e−06
2.1110	5.1e−05	2.7460	3.5e−05	1.6220	0.0234	1.6930	7.1e−05
0.7050	0.0016	0.5680	2.5e−05	0.8750	0.1784	0.6500	0.0028
0.5910	0.0002	0.3930	<1.0e−07	0.8940	0.1565	0.4400	4.0e−07
1.0890	0.6622	0.9510	0.8500	1.2460	0.2320	0.7640	0.3194
0.4710	0.0017	0.2370	1.0e−07	0.9310	0.5552	0.2550	1.2e−05

(Continued on the following page)

Table 1. Expression and *P* values for genes of interest (Cont'd)

Unigene	Gene name	Description	Map
D. Genes affected by Wnt signaling in human colorectal cancer			
Hs.127337	*AXIN2*	Axin2	17q23-q24
Hs.1578	*BIRC5*	Baculoviral IAP repeat-containing 5 (survivin)	17q25
Hs.306278	*CD44**	CD44 antigen	11p13
Hs.7327	*CLDN1*	Claudin	3q28-q29
Hs.77432	*EGFR**	Epidermal growth factor receptor	7p12
Hs.125124	*EPHB2*	EphB2	1p36.1-p35
Hs.437008	*EPHB4*	EphB4	7q22
Hs.44865	*LEF1*	Lymphoid enhancer binding factor 1	4q23-q25
Hs.7912	*NRCAM*	Neuronal cell adhesion molecule	7q31.1-q31.2
Hs.2316	*SOX9*	SRY (sex determining region Y)-box 9	17q24.3-q25.1
Hs.371279	*TCF1*	Transcription factor 1	12q22-qter
Hs.73793	*VEGF**	Vascular endothelial growth factor	6p12
E. Genes reported to be affected in human colorectal cancer			
Hs.512765	*AXIN1*	Axin1	16p13.3
Hs.79241	*BCL2*	B-cell CLL/lymphoma 2	18q21.33
Hs.1578	*BIRC5*	Baculoviral IAP repeat-containing 5 (survivin)	17q25
Hs.306278	*CD44**	CD44 antigen	11p13
Hs.194657	*CDH1*	Cadherin 1, type 1, E-cadherin (epithelial)	16q22.1
Hs.194657	*CDH1*	Cadherin 1, type 1, E-cadherin (epithelial)	16q22.1
Hs.370771	*CDKN1A*	Cyclin-dependent kinase inhibitor 1A (p21, Cip1)	6p21.2
Hs.238990	*CDKN1B*	Cyclin-dependent kinase inhibitor 1B (p27, Kip1)	12p13.1-p12
Hs.410086	*CTNNB1*	Catenin (cadherin-associated protein), β1, 88 kDa	3p21
Hs.410086	*CTNNB1**	Catenin (cadherin-associated protein), β1, 88 kDa	3p21
Hs.410086	*CTNNB1*	Catenin (cadherin-associated protein), β1, 88 kDa	3p21
Hs.82407	*CXCL16*	Chemokine (C-X-C motif) ligand 16	17p13
Hs.172648	*DLX4*	Distal-less homeobox 4	17q21.33
Hs.1602	*DPYD*	Dihydropyrimidine dehydrogenase	1p22
Hs.77432	*EGFR**	Epidermal growth factor receptor	7p12
Hs.446352	*ERBB2*	V-erb-b2 erythroblastic leukemia viral oncogene homologue 2	17q11.2-q12
Hs.169744	*G22P1*	Thyroid autoantigen (Ku70)	22q13.2-q13.31
Hs.234896	*GMNN*	Geminin, DNA replication inhibitor	6p22.1
Hs.132625	*HSF1*	Heat shock transcription factor 1	8q24.3
Hs.76473	*IGF2R*	Insulin-like growth factor 2 receptor	6q26
Hs.155995	*KIAA0643*	KIAA0643 protein	16p13.3
Hs.4276	*KIAA1701*	KIAA1701 protein	Xq23
Hs.447488	*KRAS2*	v-Ki-ras2 Kirsten rat sarcoma 2 viral oncogene homologue	12p12.1
Hs.447488	*KRAS2*	v-Ki-ras2 Kirsten rat sarcoma 2 viral oncogene homologue	12p12.1
Hs.445201	*L1CAM*	L1 cell adhesion molecule	Xq28
Hs.102267	*LOX*	Lysyl oxidase	5q23.2
Hs.110741	*MADH2*	MAD, mothers against decapentaplegic homologue 2 (*Drosophila*)	18q21.1
Hs.75862	*MADH4*	MAD, mothers against decapentaplegic homologue 4 (*Drosophila*)	18q21.1
Hs.90598	*MICA*	MHC class I polypeptide-related sequence A	6p21.3
Hs.80976	*MKI67*	Antigen identified by monoclonal antibody Ki-67	10q25-qter
Hs.433618	*MLH1*	MutL homologue 1, colon cancer, nonpolyposis type 2 (*Escherichia coli*)	3p21.3
Hs.83169	*MMP1*	Matrix metalloproteinase 1 (interstitial collagenase)	11q22.3
Hs.440394	*MSH2*	MutS homologue 2, colon cancer, nonpolyposis type 1 (*E. coli*)	2p22-p21
Hs.445052	*MSH6*	MutS homologue 6 (*E. coli*)	2p16
Hs.202453	*MYC*	V-myc myelocytomatosis viral oncogene homologue (avian)	8q24.12-q24.13
Hs.78996	*PCNA*	Proliferating cell nuclear antigen	20pter-p12
Hs.77274	*PLAU*	Plasminogen activator, urokinase	10q24
Hs.253309	*PTEN*	Phosphatase and tensin homologue	10q23.3
Hs.2090	*PTGER2*	Prostaglandin E receptor 2 (subtype EP2), 53 kDa	14q22
Hs.196384	*PTGS2*	Prostaglandin-endoperoxide synthase 2 (cyclooxygenase 2)	1q25.2-q25.3
Hs.43666	*PTP4A3*	Protein tyrosine phosphatase type IVA, member 3	8q24.3
Hs.413812	*RAC1*	Ras-related C3 botulinum toxin substrate 1	7p22

Table 1. Expression and *P* values for genes of interest (Cont'd)

Avg tumor/avg mucosa	*P*	Avg tumor/avg $Mucosa_0$	*P*	Avg tumor/avg $Mucosa_1$	*P*	Avg $Mucosa_1$/avg $Mucosa_0$	*P*
1.6360	0.0177			2.2111	6.7e−06	0.546	1.7e−06
1.5640	5.3e−06			1.8398	4.0e−07	0.723	0.0006
2.2509	2.6e−06	2.5810	4.2e−05				
4.3309	1.1e−05			8.3099	<1.0e−07		
0.7250	1.6e−04						
2.7819	5.6e−06	5.1850	<1.0e−07			3.267	1.0e−06
1.8249	1.0e−07	1.9450	3.1e−06	1.7139	5.6e−05		
0.7547	0.0007	0.7140	4.9e−04				
1.2317	0.5524						
2.2866	1.2e−05	3.2360	1.2e−06			2.001	4.3e−06
0.9190	0.8198	3.1130	1.2e−06				
2.5261	<1.0e−07	2.6250	1.0e−07	2.4320	3.0e−07		
1.2776	0.0286	1.5919	0.0005	1.0248	0.7954	1.5534	0.0055
0.3840	0.0070	0.8820	0.7758	0.2150	0.0001	4.1102	9.1e−06
1.5640	5.3e−06	1.3295	0.0087	1.8398	4.0e−07	0.7226	0.0006
2.2510	2.6e−06	2.5810	4.2e−05	1.9630	0.0019	1.3148	0.0025
0.6888	0.0368	0.9521	0.8159	0.4972	0.0018	1.9148	1.4e−05
0.6890	0.0368	0.9520	0.8159	0.4970	0.0018	3.7037	1.4e−05
0.8207	0.0293	0.8521	0.1817	0.7930	0.0226	1.0746	0.5113
1.9650	0.0001	3.0310	1.0e−07	1.2750	0.0560	2.3776	1.5e−05
0.9305	0.7134	1.6182	0.0605	0.6680	0.0382	2.4223	0.0005
1.2490	0.0624	1.6180	0.0605	1.3490	0.0364	0.6845	0.0005
0.9300	0.7134	1.1560	0.3366	0.6680	0.0382	0.8571	0.2728
0.6459	0.0211	0.3991	4.4e−05	1.0451	0.8056	0.3819	1.2e−06
0.6803	0.0034	0.4754	1.4e−06	0.9739	0.7792	0.4882	2.4e−06
0.5916	0.0037	0.6374	0.0594	0.5490	0.0117	1.1610	0.2624
0.7250	0.0002	0.7190	0.0032	0.7310	0.0035	0.9841	0.8315
0.8516	0.0783	0.7097	0.0015	1.0194	0.8329	0.6962	0.0022
1.7222	0.0014	2.9970	<1.0e−07	0.9900	0.9176	3.0266	<1.0e−07
1.8135	0.0023	2.9830	1.8e−06	1.1022	0.4607	2.7063	7.5e−05
1.0260	0.7873	0.9583	0.7322	1.0990	0.4205	0.8720	0.1660
0.7700	0.0628	0.9485	0.7610	0.6253	0.0058	1.5169	0.0018
0.9152	0.5006	0.8938	0.5417	0.9366	0.7154	0.9543	0.4934
1.4955	0.1145	1.5551	0.2377	1.4494	0.2366	1.0729	0.7143
0.7260	0.0110	1.3460	0.0949	0.5980	0.0001	1.4721	0.0168
1.2370	0.1136	0.8800	0.4033	1.1370	0.4129	1.1838	0.2708
1.2097	0.2724	1.1995	0.4985	1.2123	0.2510	0.9895	0.9618
1.7402	0.0807	1.6922	0.2840	1.7620	0.0918	0.9604	0.9362
1.0130	0.8992	1.1180	0.3950	0.9170	0.3806	1.2199	0.1393
0.3700	0.0052	0.6360	0.4052	0.2980	0.0026	2.1352	0.0116
0.8712	0.3220	0.5717	0.0002	1.3260	0.0004	0.4311	2.1e−05
1.1930	0.0152	1.2040	0.0622	1.1840	0.0534	1.0165	0.8293
0.5979	<1.e−07	0.6221	2.2e−06	0.5750	<1.0e−07	1.0819	0.3187
5.5810	<1.0e−07	4.4650	2.9e−06	6.9750	<1.0e−07	0.6401	0.0791
1.9975	0.0005	3.4692	<1.0e−07	1.1503	0.3526	3.0160	2.0e−07
1.3444	0.0003	1.4605	6.9e−05	1.2378	0.0211	1.1799	0.1342
2.1106	5.1e−05	2.7464	3.5e−05	1.6220	0.0234	1.6933	7.1e−05
2.3590	4.8e−05	3.9480	1.0e−07	1.4090	0.0757	2.8016	<1.0e−07
3.7090	<1.0e−07	3.6280	7.0e−07	3.7930	1.4e−06	0.9566	0.7533
1.0630	0.4609	1.2148	0.0671	0.9296	0.3552	1.3068	0.0206
0.6534	0.0714	0.5340	0.0429	0.7695	0.3591	0.6940	0.0488
17.7462	<1.0e−07	21.3564	3.8e−05	15.8795	1.2e−06	1.3449	0.5563
0.7080	0.1186	0.5800	0.0665	0.8450	0.5313	0.6863	0.0047
1.2145	0.0190	1.3604	0.0047	1.0850	0.3604	1.2538	0.0195

(Continued on the following page)

Table 1. Expression and *P* values for genes of interest (Cont'd)

Unigene	Gene name	Description	Map
Hs.170019	*RUNX3*	Runt-related transcription factor 3	1p36
Hs.75716	*SERPINB2*	Serine (or cysteine) proteinase inhibitor, clade B, member 2	18q21.3
Hs.7306	*SFRP1*	Secreted frizzled-related protein 1	8p12-p11.1
Hs.48029	*SNAI1*	Snail homologue 1 (*Drosophila*)	20q13.1-q13.2
Hs.360174	*SNAI2*	Snail homologue 2 (*Drosophila*)	8q11
Hs.1103	*TGFB1**	Transforming growth factor, β1	19q13.2
Hs.421496	*TGFBI*	Transforming growth factor, β-induced, 68 kDa	5q31
Hs.28005	*TGFBR1*	Transforming growth factor, β receptor I	9q22
Hs.82028	*TGFBR2*	Transforming growth factor, β receptor II (70/80 kDa)	3p22
Hs.408312	*TP53*	Tumor protein p53 (Li-Fraumeni syndrome)	17p13.1
Hs.87491	*TYMS*	Thymidylate synthetase	18p11.32
Hs.73793	*VEGF**	Vascular endothelial growth factor	6p12
Hs.248164	*WNT1*	Wingless-type MMTV integration site family, member 1	12q13
Hs.257082	*XRCC5*	X-ray repair cross-complementing group 5 (Ku80)	2q35
Hs.155040	*ZNF217*	Zinc finger protein 217	20q13.2
Hs.251664		CDNA clone MGC:52263 IMAGE:4123447, complete cds	

NOTE: A, Genes with a >5-fold change in expression and $P < 1e{-}7$ between 17 tumors and 20 mucosa as well as a >2-fold change in expression between any tumors and its patient-matched normal mucosa. B, Twenty-four genes with a >5-fold expression change were placed into two networks by the software IPA (Fig. 1). Expression and *P* values for other genes in these two networks are provided. C, Genes involved in the canonical Wnt/β-catenin pathway were examined for their differences in expression levels. D-E, Additional genes whose expression levels are reportedly affected by signaling through the Wnt/β-catenin pathway (D) or via other mechanisms (E) in human colorectal cancer. Avg tumor/avg mucosa, comparison of 17 carcinomas with 20 mucosa; Avg tumor/avg $Mucosa_0$, comparison of 17 carcinomas with 10 $Mucosa_0$; Avg tumor/avg $Mucosa_1$, comparison of 17 carcinomas with 10 $Mucosa_1$; Avg $mucosa_1$/avg $Mucosa_0$, comparison of 10 $Mucosa_1$ with 10 $Mucosa_0$.

*Genes were annotated with UniGene (all other genes were annotated with the RefSeq build September 2005).

†Oligonucleotide also corresponds to a pseudogene on chromosome 10.

‡Oligonucleotide only corresponds to splice variant 1 (NM_006238.2).

Spots with a size of <10 μm or intensity <100 in both the red and green channels were eliminated followed by removal of genes with <50% of available data. This *a priori* filtering to remove genes with unreliable signals resulted in a final tally of 12,291 genes. Of these remaining features, only those with corresponding chromosome band mapping positions (10,658) were used for calculations of average gene expression per chromosome arm.

Statistical analysis. Statistical analyses were done using the *BRBArray-Tools* package (version 3.1.0) for microarray analysis developed at the Biometric Research Branch of the National Cancer Institute (9) and MATLAB (version 6.5) from The Mathworks. For each biopsy, two technical repeat hybridizations were done and the coefficient of correlation was determined to be >90%. The replicates for each patient were therefore averaged for analysis.

A class comparison analysis was done using 17 tumors and 20 normal mucosa samples. The two-sample *t* statistic with randomized variance (10) was used to measure the difference in gene expression between the two classes. The randomized variance model assumes that the variance of the expression of each gene is randomly drawn from an inverse γ distribution and enables sharing of variance information among genes without assuming that all genes have the same variance. The stringent statistical significance threshold of $P < 0.0001$ used for the identification of genes differentially expressed between tumor and normal samples was chosen to minimize the number of false positives.

A class prediction analysis was done using the Diagonal Linear Discriminant classifier (11). Leave-one-out cross-validation (LOOCV) was used to estimate the prediction accuracy as previously described (8). After omitting one sample, the remaining samples were used to find features differentially expressed between tumor and mucosa. Genes with a $P < 0.0001$, using the two-sample *t* statistic with randomized variance, were selected. A classifier was trained using the selected genes and then used to classify the left-out sample. The average of the LOOCV error was used as an estimate of the true error. The significance of the classification results was calculated by permuting the class labels of the samples and then finding the fraction of times this relabeling resulted in higher LOOCV classification accuracy. We permuted 2,000 times, which resulted in $P < 5e{-}4$. This method takes into account the sample size and is suitable for small sample numbers.

Comparative genomic hybridization. Comparative genomic hybridization (CGH) analysis of 21 rectal carcinomas was done as previously reported (8) and can be found at http://www.riedlab.nci.nih.gov/protocols.asp. For comparison of genomic copy number and gene expression, the 500 ratio measurements per chromosome of tumor/reference calculated by Leica CW4000 imaging and analysis software (Leica, Cambridge, United Kingdom) were equally distributed over the total number of base pairs for each chromosome and plotted using Excel (Microsoft, Redmond, WA). The average genomic copy number of each chromosome arm was then calculated using the data points corresponding to each arm, excluding values that mapped to the centromeric and pericentromeric heterochromatic regions. Values were also not calculated for those chromosome arms demonstrating partial gains or losses. Average ratio values for the p-arm of the acrocentric chromosomes 13, 14, 15, 21, and 22 and the entire Y chromosome were not determined.

Biological pathway analysis. Gene lists were assessed for known biological interactions and involvement in canonical pathways using Pathways Analysis (Ingenuity, Mountain View, CA).

Results

Genes differentially expressed between normal mucosa and rectal carcinomas. Gene expression profiles from 17 tumor

Table 1. Expression and *P* values for genes of interest (Cont'd)

Avg tumor/avg mucosa	*P*	Avg tumor/avg $Mucosa_0$	*P*	Avg tumor/avg $Mucosa_1$	*P*	Avg $Mucosa_1$/avg $Mucosa_0$	*P*
1.1310	0.3436	1.0621	0.7031	1.2028	0.2550	0.8830	0.4244
1.2900	0.1839	1.3720	0.2314	1.2120	0.4452	1.1325	0.3334
0.4622	0.0028	0.2516	1.3e−05	0.8485	0.4188	0.2965	0.0001
3.7073	5.7e−05	5.5632	0.0002	2.5728	0.0003	2.1623	0.0736
1.3311	0.0733	1.6363	0.0135	1.0833	0.6657	1.5104	0.0248
0.9970	0.9850	0.7440	0.1668	1.3360	0.1633	0.5564	1.9e−06
4.3595	<1.0e−07	3.9150	<1.0e−07	4.8562	<1.0e−07	0.8062	0.1108
2.0550	5.8e−05	3.3990	<1.0e−07	1.2430	0.1093	2.7341	<1.0e−07
1.5105	0.0008	1.9947	9.0e−07	1.1421	0.2092	1.7465	0.0003
0.9618	0.6734	0.8766	0.2863	1.0558	0.6512	0.8302	0.0104
0.4944	6.9e−06	0.3385	<1.0e−07	0.7225	6.0e−05	0.4686	0.0002
2.5261	<1.0e−07	2.6247	1.0e−07	2.4320	3.0e−07	1.0793	0.3109
0.4706	0.0017	0.2373	1.0e−07	0.9314	0.5552	0.2547	1.2e−05
1.5980	0.0023	2.6190	<1.0e−07	0.9750	0.7981	2.6854	<1.0e−07
2.3257	0.0060	5.2158	3.7e−06	1.0373	0.8801	5.0282	5.5e−06
1.1372	0.2195	1.0049	0.9714	1.2871	0.0659	0.7807	0.0033

biopsies and 20 normal mucosa samples were generated by hybridization to oligonucleotide arrays (Supplementary Table S2). Of the 22,231 features on the array, 12,291 passed the filtering criteria (see Materials and Methods). These data were then used to determine the statistical significance of the differences in global gene expression. There were 1,722 genes differentially expressed at $P < 0.0001$ (using the two-sample t statistic) between the group of tumors and the normal mucosa samples. The expected number of false-positive genes, out of the 12,291 genes that passed the filtering criteria, is 1.23. Thus, the false-discovery rate was 1.23/1,722 = 0.07%. Using the even more stringent Bonferroni criteria ($P < 1.0e{-}7$), we still found 351 differentially regulated genes between the normal mucosa and rectal carcinoma groups (Supplementary Table S3). One hundred forty-four of these genes (41%) were expressed at higher levels in normal mucosa, whereas the remaining 207 genes (59%) were up-regulated in the carcinomas.

A total of 52 genes had an average expression >5-fold higher in the tumor group compared with normal mucosa, whereas only about half as many genes were expressed on average >5-fold higher in the mucosa samples (Supplementary Table S4).

To increase our confidence about the genes we were identifying as being different in expression between tumors and normal mucosa, we also looked for genes that were always >2-fold different between any tumor and its matched mucosa. We identified a group of 39 genes whose expression was always increased >2-fold in any tumor and 46 genes whose expression was always decreased >2-fold in any tumor relative to their patient-matched normal mucosa (Supplementary Table S5).

Twelve genes were found to always have a >2-fold separation in expression between tumor and its matched mucosa in addition to significant differential expression measured by the t statistic ($P < 1.0e{-}7$) and a >5-fold difference in average expression between the mucosa and tumor groups (Table 1). An additional 22 genes with significant differential expression and >5-fold difference were excluded solely on the basis that their expression difference was <2-fold in only 1 of the 12 matched pairs. Interestingly, for 18 of these 22 genes, the patient with anomalous expression was P13. The explanation for this is unknown, although one could speculate that perhaps this tumor biopsy contained a greater percentage of contaminating normal tissue.

Class prediction using global gene expression profiles. Gene expression profiles from the same set of 17 tumor biopsies and 20 normal mucosa samples were used to determine if the gene expression profile of a sample was solely sufficient to classify it as either normal rectal tissue or a rectal adenocarcinoma. This was achieved by using an established LOOCV. The number of genes used to classify the omitted sample ranged from 1,464 to 1,866, with the average being 1,611. Class prediction of the pathologic state of the assayed tissue sample was 100% accurate using a Diagonal Linear Discriminant Analysis (Supplementary Table S6).

Genes involved in rectal tumorigenesis. We used Ingenuity Pathways Analysis (IPA) to evaluate the differentially expressed genes. From the list of 351 genes with a significance value of <1.0e−7, 157 genes were identified in the knowledge bank (termed focus genes) and mapped to 29 different networks. [It must be noted for clarity that the term "network" in IPA is not the same as a biological or canonical pathway with a distinct function (i.e., cell cycle arrest) but a reflection of all interactions of a given protein as defined in the literature. Thus, both the cell cycle arrest and apoptotic pathways signaled by p53 would be part of the same network.] Nine of the networks each contained >12 genes from the list, whereas the remaining 20 networks each contained only one focus gene. The top cellular categories of the proteins in these first nine networks were cancer (58 genes), cell cycle (49 genes), cell death (62 genes), cellular movement (26 genes), cell growth and proliferation (67 genes), and tumor morphology (20 genes). For the group of 77 genes with a >5-fold change, 43 focus genes were found in the knowledge bank and two main networks were defined. The top cellular categories for the first network (consisting of 15 focus genes) were cancer, cellular growth and proliferation, and cell cycle, whereas for the second network (nine focus genes) they were immune response, tissue development, and cellular movement. These networks were connected by the expression of two genes, *prostaglandin-endoperoxide synthase 2* (*PTGS2*), also known as

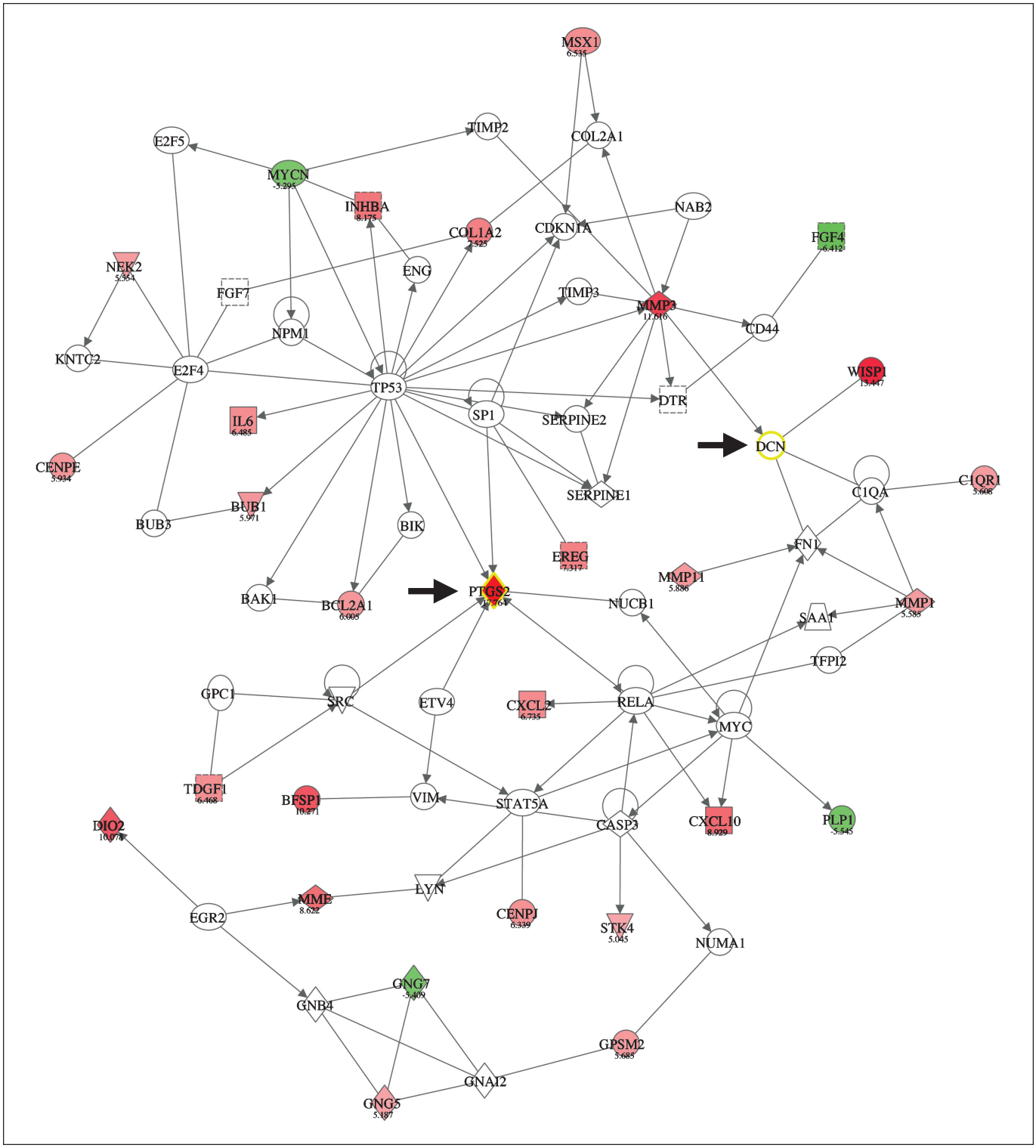


Figure 1. Network mapping of genes with >5-fold expression change using IPA. *PTGS2* and *DCN* (*arrows*) connect the top two networks identified by IPA. *Shades of red,* genes with >5-fold higher expression in the tumors; *shades of green,* >5-fold decrease in expression in the tumors relative to the normal rectal mucosa. Gene names and the actual fold changes are indicated.

cyclooxygenase 2 (*COX-2*), and *decorin* (*DCN*), only the former of which had a >5-fold difference in expression (Fig. 1). *DCN* has an average increase in expression of 2.43 (P = 2.7e−5). Nine other networks were identified, each containing only one focus gene, but many of these remaining genes fell into the cellular categories listed above.

Because Wnt/β-catenin signaling plays a prominent role in the development of colorectal cancers (5, 12, 13), we were interested in

how the expression levels of genes in this pathway were affected in rectal carcinomas. Eighteen genes on our array platform involved in the canonical Wnt/β-catenin signaling pathway were evaluated (Table 1). Only *glycogen synthase kinase 3β* (*GSK3B*) and c-*myc* (*MYC*) showed significant differences ($P < 0.0001$), both with >2-fold higher expression in the tumors. We also looked at the expression levels of 12 genes for which we found a connection between their expression level and Wnt signaling in colorectal cancer in the literature (Table 1). Eight of these genes were significantly ($P < 0.0001$) deregulated in our data set as well. Of an additional 58 genes whose altered expression was reported in colorectal cancer, we found only 13 to be differentially expressed between the tumors and the group of normal mucosa with a significance of $P < 0.0001$ (Table 1). These were *BIRC5, MLH1, MYC, PTGS2, SNAI1, TGFBI, TYMS, VEGF, CD44, MMP1, PCNA, PLAU*, and *TGFBR1*.

Genomic clustering of differentially expressed genes. We were curious as to whether there was a physical link between those 1,722 genes whose expression differed so greatly between rectal tumors and normal mucosa ($P < 0.0001$). Indeed, we found a disproportionately large number of differentially regulated genes mapping to chromosomes 13 and 20 (Fig. 2*A*). Differentially expressed genes on the remaining chromosomes revealed a distribution in a manner generally consistent with the number of genes from each chromosome spotted on the arrays (Fig. 2*A*). In general, between 50% and 75% of the deregulated genes ($P < 1.0e{-}7$) mapping to each chromosome were up-regulated in the tumors (Fig. 2*B*). This was approximately equal to the percentage of significantly up-regulated genes as a whole. A dramatically different distribution was observed on chromosomes 13 and 20, with ~85% of the differentially expressed genes having increased expression in the tumors. Of note, these chromosomes are among the most frequently observed aneuploid chromosomes in colorectal tumors (4, 14–16). Thus, it was intriguing that a disproportionate number of deregulated genes mapped to these two chromosomes and that the vast majority of them were up-regulated in the rectal carcinomas.

CGH reveals recurrent chromosome copy number alterations in rectal carcinomas. To identify chromosome copy number changes in our set of rectal tumors, we isolated genomic DNA and successfully did chromosome CGH on 21 of the 29 rectal adenocarcinomas (Supplementary Table S7; http://www.ncbi.nlm.nih.gov/sky/skyweb.cgi). As displayed in Fig. 3, gains of chromosome arms 13q (58%), 20q (46%), and 20p (38%) were in fact observed with the highest frequency, followed by copy number increases of 8q (33%) and 7p (29%). Chromosome arms 17p (46%),

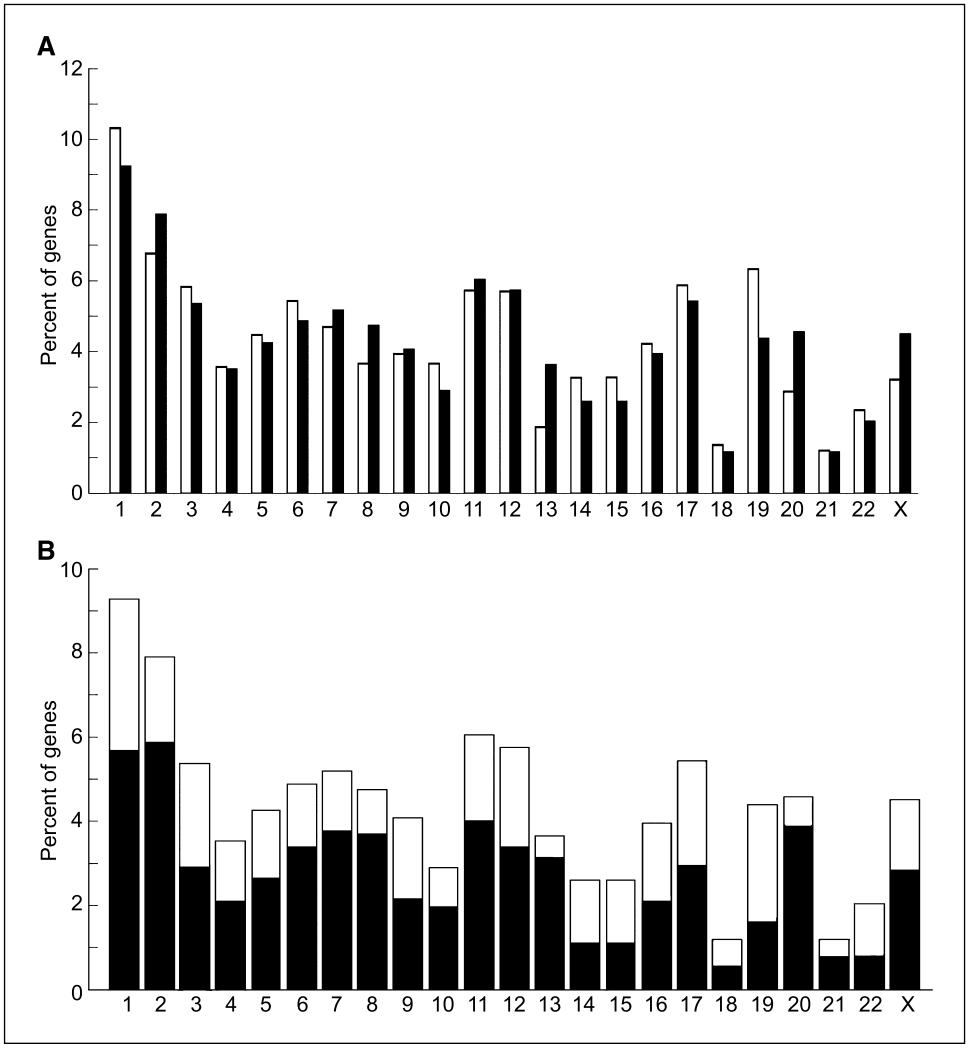


Figure 2. Chromosome localization of genes with significant expression changes. *A,* 93% of the 12,291 genes that passed the filtering criteria had chromosome mapping locations. *White columns,* percentages of these genes that map to each chromosome. Ninety-four percent of the 1,722 genes differentially expressed in the tumors with $P < 0.0001$ had known chromosome locations. *Black columns,* percentages of these genes that map to each chromosome. *B,* the percentage of genes indicated as black columns in (*A*), which were up-regulated (*black*) or down-regulated (*white*) in the tumors relative to the normal rectal mucosa.

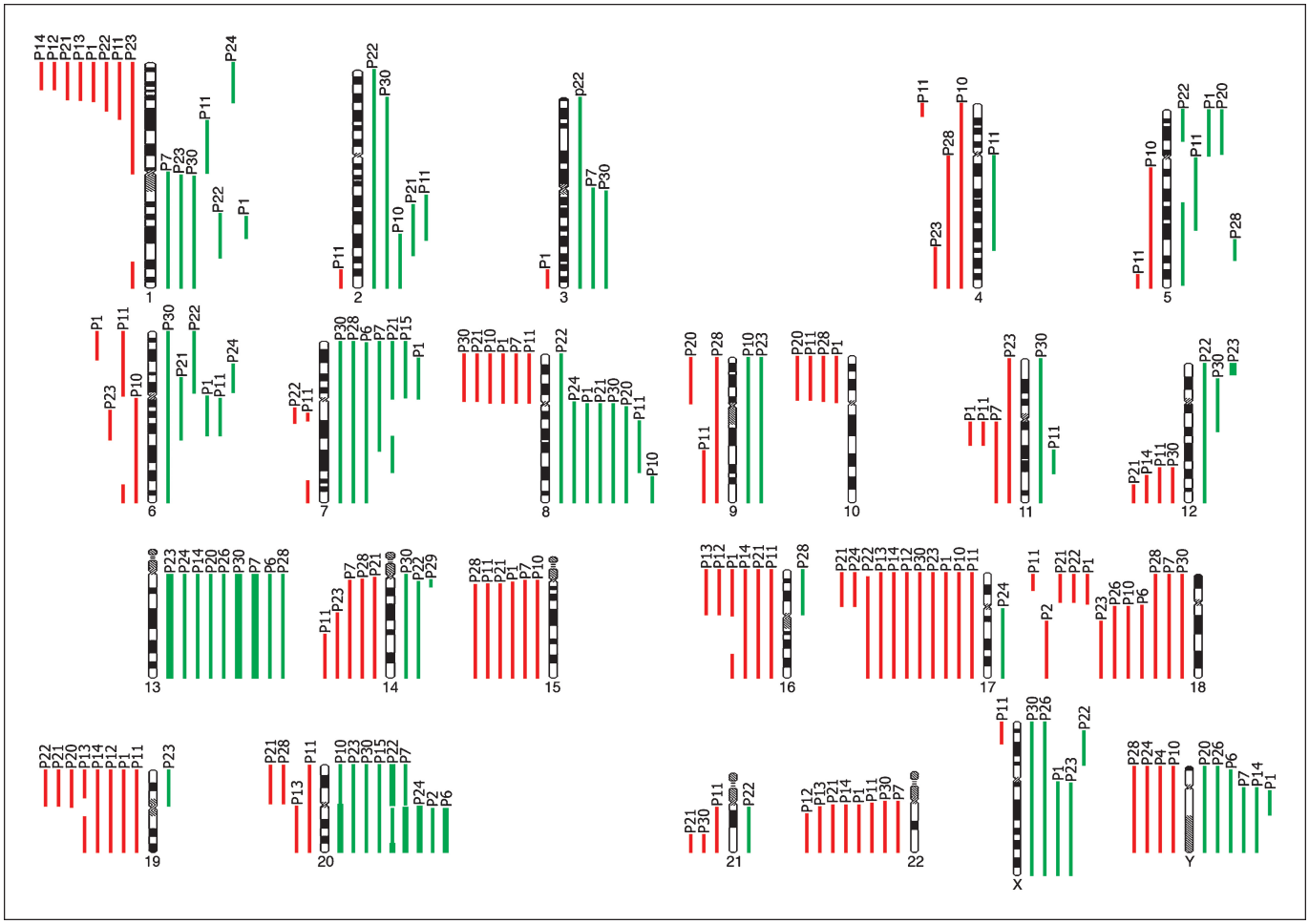


Figure 3. CGH profiles of 21 rectal carcinomas. *Red bars to the left of the chromosome,* genomic copy number losses; *green bars to the right of the chromosome,* copy number gains. *Thick bars,* high-level copy number alterations (i.e., amplifications). The patient code corresponding to each bar is indicated.

17q (38%), 19p (33%), 22q (33%), 18p (29%), 18q (29%), 8p (25%), 15q (25%), and 16p (25%) were all lost. Only three cases (P4, P17, and P28) did not display any chromosome imbalances. Another five cases (P2, P6, P12, P15, and P25) had only very few copy number changes ($\pm$ four chromosomes or chromosome arms). The remaining 13 tumors were highly aneuploid, with numerous gains and losses. P15 and P6 contained the canonical gains of chromosomes 7p, 13, and 20. P25 had a gain of 13 and loss of 18q, whereas P2 only had a gain of chromosome 20q. All of these numerical alterations have been previously associated with colorectal adenocarcinomas. Interestingly, P12 did not contain any of these normally associated aberrations, contained only chromosome losses, and could thus be considered an outlier. Of the 13 aneuploid tumors analyzed by CGH, only three contained combined gains of 7p, 13, and 20. Eight tumors contained a gain of at least one of these three chromosomes and two tumors (P11 and P13) did not have any of these whole arm gains. The latter, in fact, had a profile very similar to tumor P12.

Effects of aneuploidy on average gene expression levels. Having both the gene expression profiles and the CGH measurements, we were positioned to evaluate the extent to which these alterations were correlated. The average chromosome arm expression level (relative to the Stratagene reference RNA) was computed using 12 tumor biopsies for which we had both expression and CGH data. Similarly, we averaged the CGH values measured along the length of each chromosome arm to compute an average copy number for each chromosome arm. These values were plotted against each other for the 10 chromosome arms frequently gained or lost (7p, 8p, 8q, 13q, 17p, 17q, 18p, 18q, 20p, and 20q). We determined both the percentage correlation and the R^2 values between the average arm expression values and the average CGH measurement. In general, there was a strong positive correlation between the chromosome arm copy number and the average expression of genes localized on that arm. Results for the six arms with very strong correlations and a significance of $P < 0.05$ are illustrated in Fig. 4*A*. The remaining four arms had a weaker correlation and higher P values (Supplementary Table S8). Correlations were also calculated for the remaining 31 chromosome arms. However, these results are more difficult to interpret due to the lower frequencies (often only one case) with which they were gained or lost (Supplementary Table S8). Another way of looking at this correlation is to plot the average expression of each gene along the length of the arm as we have previously done for artificially induced trisomies (17). Figure 4*B* shows the average gene expression for those patients without any copy number alteration (*left*) and the average gene expression for those tumors with a gain (7p, 8q, 13q, 20p, and 20q) or loss (18q) of that arm (*right*). The immediate association of chromosome arm average

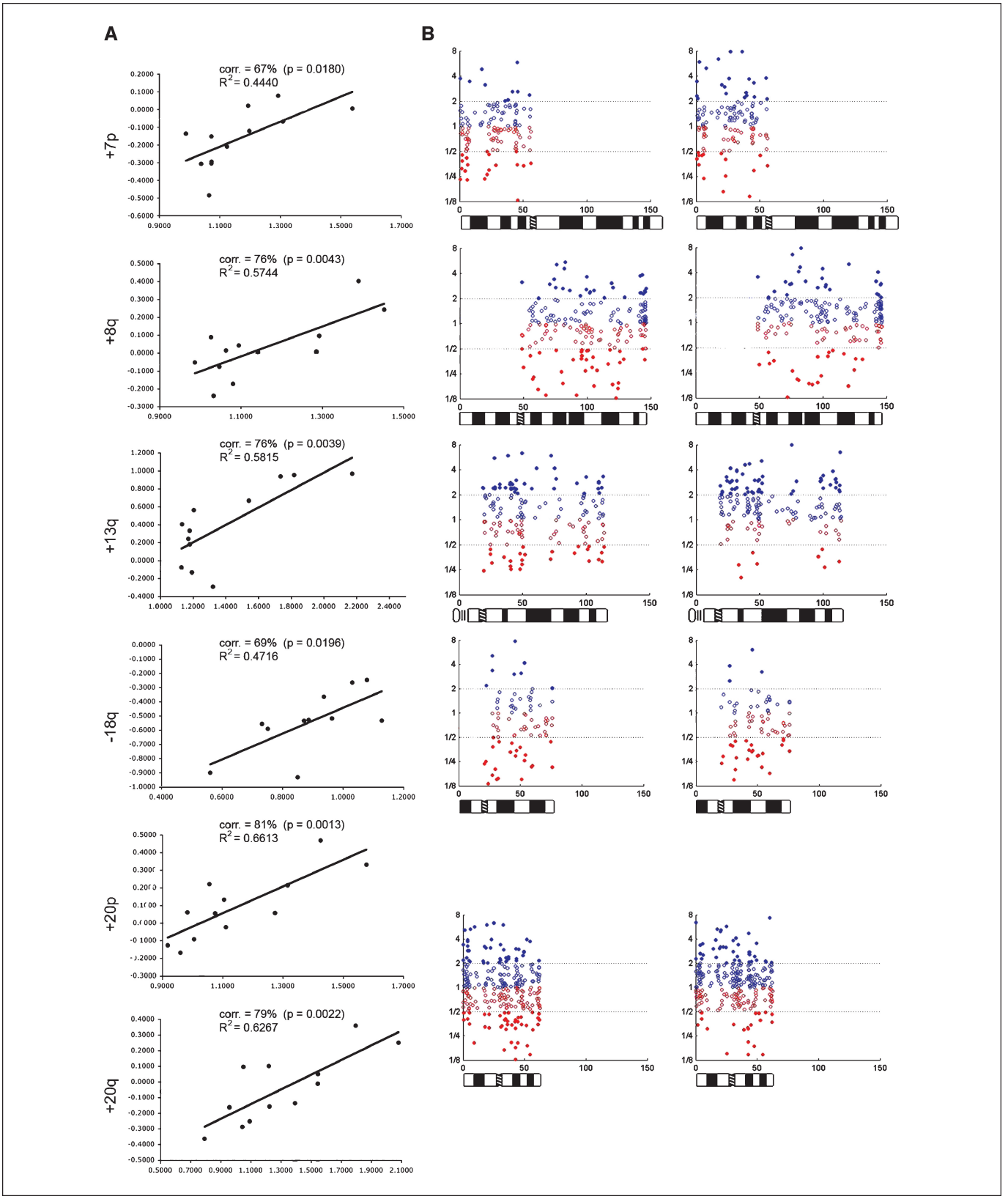
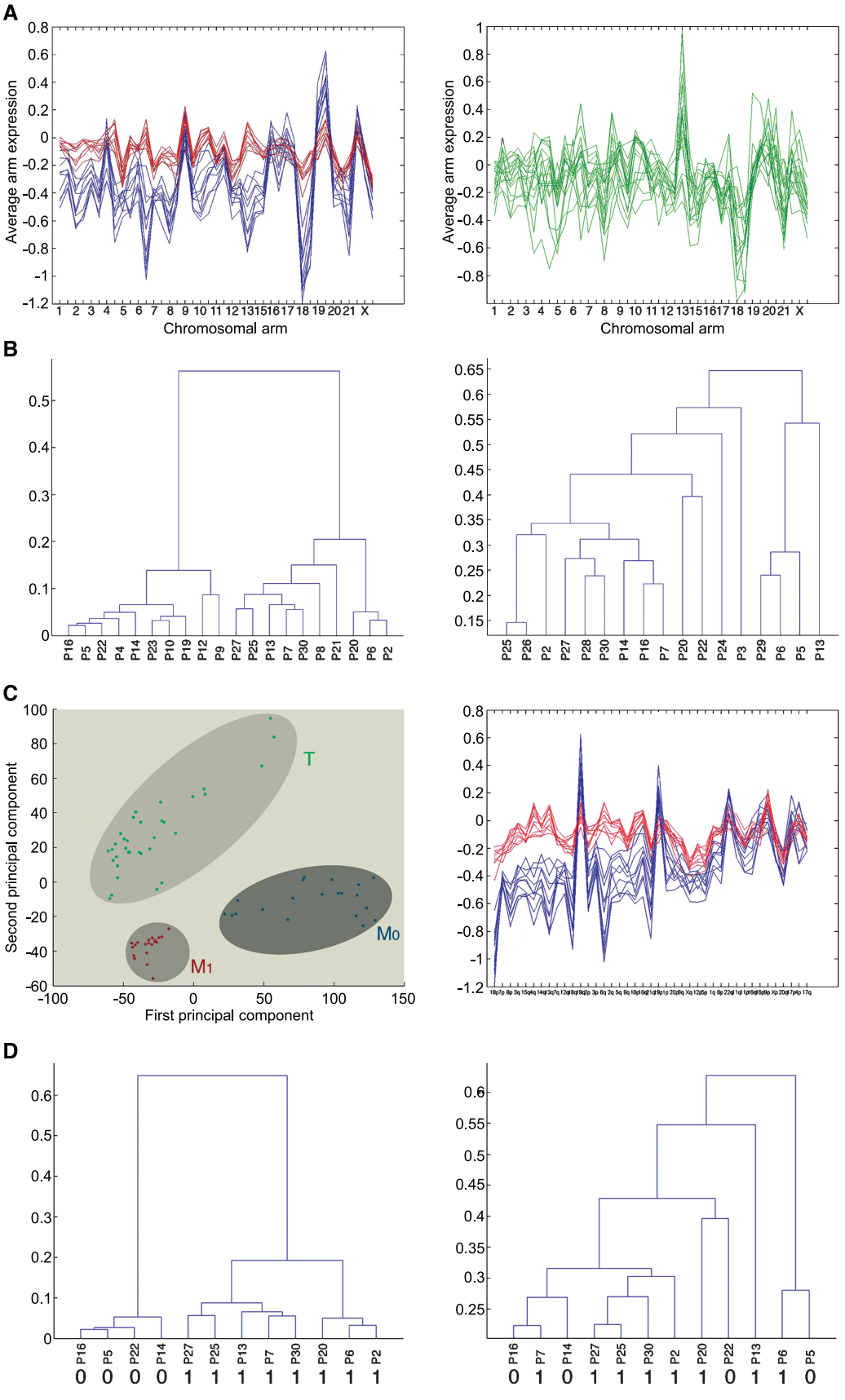


Figure 4. Correlation between gene expression and alterations of chromosome copy number. *A,* the average gene expression value (*Y* axis) is plotted against the average CGH ratio value (*X* axis) for each of 12 patients for which we had done both analyses. The percentage correlation, its *P* value, and the R^2 are indicated on each plot. The directionality of the copy number change most commonly observed is represented as a gain (+) or loss (−) preceding the chromosome number. *B,* the average expression of each gene along the length of the chromosome is plotted for those carcinomas without (*left*) and with (*right*) a copy number alteration. These plots correspond to the graphs in (*A*). Specifics about the number of samples in each category are included in Supplementary Table S8. *Blue,* genes with increased expression relative to the reference RNA; *red,* genes with decreased expression relative to the reference RNA.


A
Average arm expression
Chromosomal arm
B
C
Second principal component
First principal component
T
M0
M1
D

gene expression levels and chromosomal copy number is clearly depicted as a general shift in the expression profiles.

Chromosome arm expression reveals two normal rectal mucosa subgroups. When we plotted the chromosome arm expression profiles, inspection of the normal mucosa samples immediately revealed two distinct expression patterns (Fig. 5*A, left*). The carcinomas, however, did not display such a bipolarity (Fig. 5*A, right*). In general, chromosomes 8q, 13, and 20 displayed a tendency toward high expression, whereas chromosomes 8p, 18, and 21 showed lower average expression. Unsupervised clustering based on chromosome arm average expression values confirmed the bipolarity of the normal mucosa samples (Fig. 5*B, left*). Again, the 17 carcinomas did not partition into distinct groups (Fig. 5*B, right*). Notably, the amount of heterogeneity was higher in the carcinomas than within either of the two groups of normal mucosa. In fact, the small amount of heterogeneity within one of the mucosa groups ($Mucosa_1$) was quite remarkable. The principle component analysis in Fig. 5*C* (*left*) clearly reflects both the separation of mucosa and tumor biopsies into three groups and the varying amount of heterogeneity in each one. The average chromosome arm–specific expression level of 18p (with a *t* statistic of 158.1) was best suited for separation of the two normal mucosa groups, followed by the average expression level of chromosome arms 7p, 8p, 3q, 15q, 4q, 14q, and 13q, respectively (Fig. 5*C, right*). As noted above, some of these chromosomes are typically subject to copy number changes (7p, 8p, 13q, and 18p), whereas others are not (3q, 15q, 4q, and 14q).

Differences between the two normal mucosa groups. We then wished to determine the origin of these distinct normal rectal expression profiles. The most obvious possibility was that these *a priori* differences would somehow affect the pathway embarked upon by the early tumor progenitors, and this choice would be reflected in the expression profiles observed in the matched rectal carcinomas. When we did unsupervised cluster analysis of the normal mucosa from the 12 matched samples in our data set, again using the average chromosome arm expression values, the four normal $Mucosa_0$ are clearly separated from the remaining eight samples of $Mucosa_1$ (Fig. 5*D, left*). When we did the same analysis on the matched tumor biopsies, however, the patients did not partition into two distinct branches (Fig. 5*D, right*). In fact, the most divergent tumors in terms of their expression profiles (P16 and P5) came from normal mucosa with the least profile divergence.

Because using the more global average arm expression values was uninformative in determining the reason for the bifurcation of the normal mucosa samples, we looked at the differences between the tumors and the two groups of normal mucosa at the level of individual genes. Consistent with dendrograms showing that normal $Mucosa_1$ was closer to the tumor samples than $Mucosa_0$, a much larger set of genes was differentially expressed between the tumors and $Mucosa_0$ ($\sim$3,700 with $P < 0.0001$ and 792 with $P <$ 1.0e−7) than between tumor and $Mucosa_1$ (1,066 with $P < 0.0001$ and 112 with $P <$ 1.0e−7). Of those genes with $P <$ 1.0e−7, 49 were significantly different between the tumors and both groups of mucosa. When we looked at 20 other genes in the canonical Wnt signaling pathway (Table 1), only one gene (*GSK3B*) was significantly different between the tumors and the normal $Mucosa_1$, whereas 12 genes were significantly different between the tumors and $Mucosa_0$. There were also nine genes, including *WNT1* itself, with significant differential expression between the two groups of normal mucosa. Likewise, of the 58 differentially expressed genes previously identified in colorectal cancer (Table 1), we found 23 genes significantly different between tumors and $Mucosa_0$ and eight between tumors and $Mucosa_1$.

When we then looked to see what distinguished the two normal mucosa groups one from another, we found 574 genes differentially expressed at $P <$ 1.0e−7, 134 of which had an average fold difference >5 (Supplementary Table S9). Of interest were many genes involved in DNA damage signaling and repair [*ATR, G22P1* (Ku70), *MRE11A, TSN*, and *XRCC5* (Ku80)], proteins involved in the ubiquitin and SUMO degradation pathways (USP1, USP3, and UBA2), and histone/histone modification enzymes (DOT1L, HDAC1, and H2AV). Additional genes in these categories were also differentially expressed, albeit at a slightly lower *P* value. Sixteen of the colorectal cancer–associated genes were significantly deregulated between the two categories of normal mucosa. Some of these encode proteins involved in DNA double-strand break [Ku70 (*G22P1*) and Ku80 (*XRCC5*)] and mismatch repair (*MSH2*), cell cycle (*CDKN1B*), growth (*MYC, PCNA1, TGFB1*, and *TGFBR1*) and survival (*BCL2*; Table 1).

Discussion

The overall pattern of chromosomal copy number alterations we observed in 21 rectal carcinomas was mainly consistent with previous analyses of colorectal carcinomas (4, 14–16). In particular, gains frequently involved chromosome arms 7p, 8q, 13q, 20p, and 20q, whereas losses often affected 8p, 17p, 17q, 18p, and 18q. We also identified a decreased copy number of chromosomes 15, 16p, 19p, and 22q in at least one quarter of the rectal tumors. Therefore, there do not seem to be cytogenetically appreciable differences in the chromosomes preferentially gained or lost in colon versus rectal carcinomas. It will be very interesting to see if they are as similar at the level of individual gene expression patterns. Such analyses are currently being conducted.

We and others have previously explored the relationship between chromosomal aneuploidy and average gene expression levels in a number of different biological systems. One example involved the creation of artificial trisomies in karyotypically normal cell lines (17). Comparison of expression profiles of the derived and parental cell lines showed a clear and immediate effect, as the average gene expression levels increased for the trisomic chromosomes. A similar picture emerged from four other reports of primary breast cancers and breast cancer–derived cell lines analyzed by CGH and expression profiling. These studies, looking at specific amplicons, also suggest a direct

Figure 5. Average chromosome arm expression patterns clearly discern tumors from rectal mucosa but also reveal two categories of normal mucosa. *A,* the gene expression values for every gene along the length of an individual chromosome arm were averaged for each of 20 normal mucosa biopsies (*left*) and 17 rectal carcinomas (*right*). There was a visible separation of the mucosa into two groups, each composed of 10 biopsies. The groups are differentially colored, representing $Mucosa_0$ (*blue*) and $Mucosa_1$ (*red*). *B,* unsupervised hierarchical cluster analysis of 20 normal mucosa (*left*) and 17 rectal carcinomas (*right*) using the average chromosome arm gene expression values plotted in (*A*). *C,* principle component analysis (*left*) based on the average chromosome arm gene expression values. *Right,* plot of the average chromosome arm gene expression values for the 20 normal mucosa from left to right, which are most useful in distinguishing $Mucosa_0$ (*blue*) from $Mucosa_1$ (*red*). *D,* unsupervised hierarchical cluster analysis of the 12 patients from whom we had normal rectal mucosa (*left*) and patient matched tumor biopsies (*right*). *0,* $Mucosa_0$; *1,* $Mucosa_1$.

effect of genomic copy number changes on gene expression levels (18–22). Consistent with these data, our analysis similarly shows a strong correlation between copy number changes and the average gene expression levels for those chromosomes frequently gained or lost in rectal carcinomas. We therefore conclude that there is in fact an aneuploidy-induced deregulation of the cancer transcriptome that occurs in addition to the transcriptional and mutational deregulation of oncogenes and tumor suppressor genes.

Surprisingly, chromosome arm–specific average gene expression values also revealed a partitioning of the normal mucosa samples into two groups. For instance, average gene expression levels of chromosomes 18p, 7p, 8p, 3q, 15q, 4q, 14q, and 13q were clearly different between $Mucosa_1$ and $Mucosa_0$. No correlation could be found between the average arm expression profiles of these two normal rectal mucosa groups and any clinical attributes, e.g., the investigating surgeon (sample ascertainment bias), the location of the tumor or normal mucosa biopsies within the rectum, the gender of the patients, their age, history of familiar colorectal cancer, the patient's response to chemoradiotherapy, the carcinoma CGH profile, or systematic errors due to the experimental design. One possible explanation would be if $Mucosa_1$ biopsies represent "transitional mucosa," surrounding tissue whose gene expression is being directly influenced by its proximity to the tumor (23, 24). At the time of sample ascertainment, however, the exact location of the normal biopsy with regard to distance form the tumor was not recorded. One immediate outcome of this study is that this information along with histology will be noted in future procedures conducted at the Department of General Surgery at the University Medical Center, Göttingen.

The observed differences in gene expression patterns from the mucosa did not persist as differences in the tumor transcriptome. Additionally, the number of genes differentially regulated was smaller between tumor and $Mucosa_1$ than between tumor and $Mucosa_0$, implying a closer relationship. Analysis of the genes separating the mucosa groups showed an up-regulation of genes involved in response to DNA damage, histone modification, RNA posttranscriptional modification, and the ubiquitin pathway in $Mucosa_1$. Two recent reports have shown such an increase in DNA damage checkpoint genes in precancerous lesions (25, 26). One could, therefore, also speculate that $Mucosa_1$ represents a morphologically normal, but transcriptionally altered precancerous lesion, consistent with the notion of "field cancerization" (27).

Using individual gene expression profiles, we were able to correctly classify the rectal tumors and the corresponding normal rectal mucosa samples with 100% accuracy, a reflection of the large number of genes whose expression significantly differed. Notably, we identified more significantly different genes mapping to chromosomes 13 and 20 than one would expect based on the number of sequences from these chromosomes spotted on the array (Fig. 2*A*), and the ratio of up-regulated to down-regulated genes mapping to these two chromosomes was also higher than for the other chromosomes (Fig. 2*B*). One explanation could be that these were the most frequently gained chromosomes; therefore, genes mapping to them are more likely to be expressed at a higher level (Fig. 3). Alternatively, increased expression of specific genes on these chromosomes may be necessary for tumorigenesis and, hence, selection for a gain of these chromosomes may be one mechanism by which to accomplish this.

In looking through our lists of significantly deregulated genes, we were reassured to find a number of genes whose association with colorectal cancer has already been reported. Of particular note was *PTGS2,* the pharmacologic target of COX-2 inhibitors, which had an average fold increase of 17.764 ($P < 1.0e{-}7$). Consistent with previous reports, the expression level of *COX-1* (*PTGS1*) was not altered in the tumors compared with the surrounding normal mucosa (fold increase = 0.943, $P < 0.6342$). COX-2 is believed to exert its effects through an increase in the conversion of arachidonic acid to prostaglandin H2, which is then converted to, among other prostaglandins, prostaglandin E_2 (PGE_2). We therefore looked at the regulation of genes involved in some of the pathways affected by PGE_2. These included inhibition of apoptosis through increased *Bcl-2* levels (0.384, $P = 0.007$; ref. 28), increased cellular proliferation induced by signaling through epidermal growth factor receptor (*EGFR;* 0.725, $P = 0.0002$; ref. 29) and angiogenesis via increased expression of *VEGF* (2.525, $P < 1.0e{-}7$; ref. 30). Because both *Bcl-2* and *EGFR* are down-regulated in the tumor samples, we have made the assumption that the main role of *PTGS2* during tumor formation is in supplying vasculature to the tumor rather than through the prevention of apoptosis or an increase in cellular proliferation.

It has been reported that the Wnt signaling pathway can regulate *COX-2* expression (31). At odds with this observation and the wealth of data on Wnt pathway activation in colorectal cancer (32, 33), we found that the expression levels of *Wnt1* (0.4706, $P = 0.0017$) and *Wnt 3A* (0.7724, $P = 0.0187$) were actually decreased in the rectal carcinomas relative to the surrounding mucosa. *Wnt8,* the other family member known to activate the canonical Wnt/β-catenin pathway, was not spotted on our arrays and could therefore not be assessed. *GSK3B,* which is responsible for phosphorylation of β-catenin leading to its degradation, is significantly up-regulated in the rectal tumors (2.0293, $P < 1.0e{-}7$). According to a recent review (33), this would be the first report of altered *GSK3B* expression in tumors and is in fact contrary to the expected down-regulation. We cannot exclude the possibility that a kinase-inactive mutant of the protein is being overexpressed, a result that would be consistent with recent findings in a mouse model of mammary tumorigenesis (34). Altered expression of other proteins in the β-catenin degradation complex was generally less significant, but also reflect an increased expression: *Axin1* (1.277, $P = 0.0286$), *Axin2/conductin* (1.6360, $P = 0.0177$), *casein kinase 1α1* (1.6910, $P = 0.0418$), and *casein kinase 1ε* (1.6737, $P = 1e{-}5$).

Based solely on the expression levels of these proteins, it seems that contrary to current models, the tumor cells were in a position to increase degradation of β-catenin, thereby abrogating signaling through this canonical pathway. In agreement with this assessment, the expression levels of α- and *β-catenin* were not significantly altered ($P = 0.06$ to 0.71), and *δ1-catenin* (0.6976, $P = 0.0003$) and *δ2-catenin* (0.1819, $P = 1.1e{-}6$) were significantly decreased. The caveat was *catenin β–like 1* (2.2283, $P < 1.0e{-}7$), which interestingly maps to 20q11.23-q12. Transient overexpression of this protein in Chinese hamster ovary cells, however, resulted in increased apoptosis (35).

Looking at the expression level of genes whose transcriptional activity has been reported to be increased through Wnt/β-catenin signaling, we see that *PPARD* (0.5924, $P = 1.6e{-}4$), *EGFR* (0.7250, $P = 1.6e{-}4$), *TWIST2* (0.3852, $P < 1.0e{-}7$), *IGF2* (0.3125, $P < 1.0e{-}7$), *KRTBH4* (0.3731, $P < 1.0e{-}7$), and *FGF4* (0.1837,

P = 7.9e−6) are actually down-regulated, whereas *SOX9* (2.2866, P = 1.19e−5), c-*myc* (*MYC*: 2.1110, P = 5.13e−5), *survivin* (*BIRC5*: 1.5637, P = 5.3e−6), *cyclin D1* (*CCND1* 1.3893, P = 0.0078), *CD44* (2.2509, P = 2.6e−6), *Ephrin B2* (2.7819, P = 5.6e−6), *Ephrin B4* (1.8249, P = 1.0e−7), *CLDN1* (4.3309, P = 1.1e−5), *VEGF* (2.5252, P < 1.0e−7), *Dickkopf* (*DKK3*: 2.1477, P = 9.8e−6), *WISP1* (13.4472, P < 1.0e−7), *IL6* (6.4850, P < 1.0e−7), *keratin 12* (1.4661, P = 1.1e−4), *keratin 17* (3.8805, P < 1.0e−7), *keratin 23* (12.3390, P < 1.0e−7), and *PGTS2* were significantly increased. Taking all of these results into account, it seems that the β-catenin pathway was in fact activated in the rectal tumors. This could potentially be achieved through mutation of phosphorylation site serine and threonine residues of β-catenin itself, which would protect the protein from the increased presence of the β-catenin degradation complex (a relatively infrequent event occurring in ~7% of colorectal carcinomas). Another possibility is that mutation of APC, and hence one of the major components of the degradation complex, is responsible for stabilization of β-catenin. This is not unlikely given the fact that APC mutations are actually found in ~80% of sporadic colorectal cancers.

Because activation of a biological signaling pathway often leads to interaction with several different downstream affecters, our analysis of these rectal carcinomas begins to identify which portions of the Wnt signaling pathway may in fact be necessary for tumor development and which are dispensable. A more detailed look into the regulation of those genes whose transcriptional activity is contradictory to expectation might be extremely informative with respect to understanding the complex regulation of the transcriptome observed in tumor cells.

Use of the IPA software enabled us to identify interacting genes within our networks that were not part of our focus gene lists. One means of identifying false positives or genes whose altered expression may not be directly linked to tumorigenesis is to query why these interacting genes were not identified as significantly altered in their expression levels. For instance, *PTGS2* is significantly increased and forms a link between the two networks illustrated in Fig. 1. Of the genes known to regulate its expression, one is significantly increased (*ETV4*; 2.7-fold, P < 1.0e−7), one is decreased 1.75-fold (*SRC*; P = 0.0002), and two are unaffected (*TP53*, 0.9619, P = 0.6734 and *RELA*, 0.8308, P = 0.0246; Table 1). Although the expression level of TP53, the central hub of the upper network, is unaffected, many of the genes that it regulates or interacts with had >5-fold increases in expression (Supplementary Table S4: *INHBA, COL1A2, IL6, BCL2A1, BUB1, PTGS2, CENPE*, and *NEK2*). On the other hand, a number of other TP53-responsive genes were not affected to this extent. When we asked why this was the case, we saw that for some genes—although their P values were significant—their fold changes were not >5-fold (*BUB3, ENG*, and *NPM1*; Table 1). For others, it was because they were absent from our array (*E2F2, SERPINE2, KNTC3*, and *DTR*), and the remaining genes in this TP53-centered network simply had neither a significant nor a large fold change (*BAK1, BLK, SERPINE1, TIMP3, SP1, TIMP2*, and *E2F5*). *BAK1* and *BLK* induce cell death; hence, it is perhaps not surprising that these two TP53-regulated genes are not increased in expression in rectal adenocarcinomas. *TIMP2* and *E2F5* are regulated by *MYCN*, which was reduced in expression by ~8-fold (P < 1.0e−7) in the tumors. *DCN*, the other gene linking the two networks in Fig. 1, has been shown to suppress tumor cell line growth. This function is seemingly at odds with the increased expression level observed in our rectal adenocarcinomas.

In summary, we uncovered an apparent discrepancy in utilization of the Wnt/β-catenin signaling pathway with previous colorectal cancer literature. We hypothesize a bifurcation of this canonical pathway with respect to the downstream target genes in rectal carcinogenesis. We identified a series of differentially expressed genes in rectal carcinomas that have not previously been associated with colorectal tumorigenesis. Combined application of our conservative statistical criteria came down to a specific set of 12 differentially expressed genes that we feel play a crucial role in the development of rectal tumors. In addition, we saw a general deregulation of the transcriptome in the tumors. For those chromosome arms we identified by CGH to be most frequently aneuploid in the rectal carcinomas, there was a very strong correlation with the average transcriptional activity of its resident genes. These data, along with that presented in other experimental models of aneuploidy, provide additional evidence that genomic imbalances indeed affect the transcriptome and thus support a role for aneuploidy in tumorigenesis.

Acknowledgments

Received 7/25/2005; revised 10/4/2005; accepted 10/21/2005.

Grant support: Intramural Research Program of the NIH, National Cancer Institute.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Drs. Cristina Montagna, Amanda Hummon, and Quang Tri Nguyen for their invaluable advice and assistance; Vasuki Gobu and Klaus Moderegger for extracting chromosome CGH values from the Leica software; Buddy Chen for IT support and manuscript editing; and Dr. Reinhard Ebner for critical suggestions on the manuscript.

References

1. Midgley R, Kerr D. Colorectal cancer. Lancet 1999;353:391–9.
2. Bardi G, Sukhikh T, Pandis N, Fenger C, Kronborg O, Heim S. Karyotypic characterization of colorectal adenocarcinomas. Genes Chromosomes Cancer 1995;12:97–109.
3. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 1990;61:759–67.
4. Ried T, Knutzen R, Steinbeck R, et al. Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. Genes Chromosomes Cancer 1996;15:234–45.
5. Leslie A, Carey FA, Pratt NR, Steele RJ. The colorectal adenoma-carcinoma sequence. Br J Surg 2002;89:845–60.
6. Ried T, Heselmeyer-Haddad K, Blegen H, Schrock E, Auer G. Genomic changes defining the genesis, progression, and malignancy potential in solid human tumors: a phenotype/genotype correlation. Genes Chromosomes Cancer 1999;25:195–204.
7. Sauer R, Becker H, Hohenberger W, et al. Preoperative versus postoperative chemoradiotherapy for rectal cancer. N Engl J Med 2004;351:1731–40.
8. Ghadimi BM, Grade M, Difilippantonio MJ, et al. Effectiveness of gene expression profiling for response prediction of rectal adenocarcinomas to preoperative chemoradiotherapy. J Clin Oncol 2005;23:1826–38.
9. Simon R, Peng A. BRBArrayTools; 2003.
10. Wright GW, Simon RM. A random variance model for detection of differential gene expression in small microarray experiments. Bioinformatics 2003;19:2448–55.
11. Dudoit S, Fridlyand J. A prediction-based resampling method for estimating the number of clusters in a dataset. Genome Biol 2002;3:RESEARCH0036.
12. Korinek V, Barker N, Morin PJ, et al. Constitutive transcriptional activation by a β-catenin-Tcf complex in APC−/− colon carcinoma. Science 1997;275:1784–7.
13. Arends JW. Molecular interactions in the Vogelstein model of colorectal carcinoma. J Pathol 2000;190:412–6.
14. Hermsen M, Postma C, Baak J, et al. Colorectal adenoma to carcinoma progression follows multiple

pathways of chromosomal instability. Gastroenterology 2002;123:1109–19.
15. Ghadimi BM, Grade M, Liersch T, et al. Gain of chromosome 8q23-24 is a predictive marker for lymph node positivity in colorectal cancer. Clin Cancer Res 2003;9:1808–14.
16. Knosel T, Schluns K, Stein U, et al. Chromosomal alterations during lymphatic and liver metastasis formation of colorectal cancer. Neoplasia 2004; 6:23–8.
17. Upender MB, Habermann JK, McShane LM, et al. Chromosome transfer induced aneuploidy results in complex deregulation of the cellular transcriptome in immortalized and cancer cells. Cancer Res 2004;64: 6941–9.
18. Monni O, Barlund M, Mousses S, et al. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. Proc Natl Acad Sci U S A 2001;98:5711–6.
19. Virtaneva K, Wright FA, Tanner SM, et al. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. Proc Natl Acad Sci U S A 2001; 98:1124–9.
20. Pollack JR, Sorlie T, Perou CM, et al. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc Natl Acad Sci U S A 2002;99: 12963–8.
21. Hyman E, Kauraniemi P, Hautaniemi S, et al. Impact of DNA amplification on gene expression patterns in breast cancer. Cancer Res 2002;62:6240–5.
22. Wolf M, Mousses S, Hautaniemi S, et al. High-resolution analysis of gene copy number alterations in human prostate cancer using CGH on cDNA microarrays: impact of copy number on gene expression. Neoplasia 2004;6:240–7.
23. Leuenroth S, Riethdorf S, Erbersdobler A, et al. Detection of human telomerase RNA in the tumour-surrounding mucosa of bladder carcinomas as a marker for premalignant transformation. BJU Int 2005;96:553–7.
24. Wang QA, Gao H, Wang YH, Chen YL. The clinical and biological significance of the transitional mucosa adjacent to colorectal cancer. Jpn J Surg 1991;21:253–61.
25. Gorgoulis VG, Vassiliou LV, Karakaidos P, et al. Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. Nature 2005;434:907–13.
26. Bartkova J, Horejsi Z, Koed K, et al. DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. Nature 2005;434:864–70.
27. Ha PK, Califano JA. The molecular biology of mucosal field cancerization of the head and neck. Crit Rev Oral Biol Med 2003;14:363–9.
28. Sheng H, Shao J, Morrow JD, Beauchamp RD, DuBois RN. Modulation of apoptosis and Bcl-2 expression by prostaglandin E2 in human colon cancer cells. Cancer Res 1998;58:362–6.
29. Schlessinger J. Cell signaling by receptor tyrosine kinases. Cell 2000;103:211–25.
30. Cianchi F, Cortesini C, Bechi P, et al. Up-regulation of cyclooxygenase 2 gene expression correlates with tumor angiogenesis in human colorectal cancer. Gastroenterology 2001;121:1339–47.
31. Araki Y, Okamura S, Hussain SP, et al. Regulation of cyclooxygenase-2 expression by the Wnt and ras pathways. Cancer Res 2003;63:728–34.
32. Wang D, Mann JR, DuBois RN. WNT and cyclooxygenase-2 cross-talk accelerates adenoma growth. Cell Cycle 2004;3:1512–5.
33. Behrens J, Lustig B. The Wnt connection to tumorigenesis. Int J Dev Biol 2004;48:477–87.
34. Farago M, Dominguez I, Landesman-Bollag E, et al. Kinase-inactive glycogen synthase kinase 3β promotes Wnt signaling and mammary tumorigenesis. Cancer Res 2005;65:5792–801.
35. Jabbour L, Welter JF, Kollar J, Hering TM. Sequence, gene structure, and expression pattern of CTNNBL1, a minor-class intron-containing gene-evidence for a role in apoptosis. Genomics 2003;81: 292–303.